

Semantic and Social Networks Comparison for the Haiti Earthquake Relief Operations from APAN Data Sources using Lexical Link Analysis (LLA)

Dr. Ying Zhao, Dr. Shelley P. Gallup, and Dr. Douglas J. MacKinnon
Distributed Information Systems Experimentation (DISE) Group
Naval Postgraduate School, Monterey, CA 93943
yzhao@nps.edu, spgallup@nps.edu, djmackin@nps.edu

1 Abstract

The Haiti earthquake relief effort in 2010 involved numerous civil and military organizations engaged in vigorous relief operations, exemplifying rapid deployment of logistics, transport, security and medical supplies. Yet, to determine the effectiveness of these operations, and to understand how these various organizations collaborated, and eventually how to improve the future overall effectiveness of these operations, requires sifting through large volumes of communications and determining a methodology so that analysts may understand how to gain understanding and analyze such data.

We have been leveraging “Lexical Link Analysis” (LLA), as one means to achieve “System Self-awareness” (SSA), and develop this as a methodology to analyze large text-based data sets, such as the interagency collaborations and problem solving processes using documents obtained from traditional communication platforms as well as new social media. Last year, we reported using LLA to analyze open source data sources of the Haiti relief operations. This year, we will report on the analysis applied to the APAN (The All Partners Access Network) forum discussion data from the relief effort.

2 Introduction

As we introduced and reported previously (Zhao, etc., 2011), in the aftermath of the Haiti earthquake, US military and civil organizations provided rapid and extensive relief operations. Organizations involved in the operation collaborated with each other to negotiate roles, define limits, set business rules etc., to maximize efficiency and decrease response times. To determine that the expectations for effective and efficient operations had been attained would require in-depth analysis of collaboration, dialog and resultant outcomes. Such analysis by manual means would require numerous man-hours and arduous effort.

The challenge is to read and understand the data that are collected in real-life events and to gain insight on what the roles and relationships of these organizations were and how the operations were actually conducted, in terms of how various organizations (military and civil) actually interacted- and how their synergies were achieved.

In a technical sense, data sources for analyzing collaborations among military, government and civil stakeholders typically include unstructured data such as free text, word, .pdf, or PowerPoint documents in collaboration websites such as APAN, and new and varied platforms of communication such as the internet, intranet, and myriad social media. The data has been overwhelming yet may contain the critical insights and could impact and help the management of future operations. Cross-examining all this data requires automated tools which in the future may provide real-time analysis when applied to future events.

The US SOUTHERN COMMAND engaged in the HAITI Humanitarian Assistance/Disaster Relief (HA/DR) Community of Interest (COI) on the All Partners Access Network (APAN) during the Haiti crisis. The APAN data was captured from an MSSQL database. The samples from the data include:

- Official documents and briefings in PDF: ~167 PDF file attachments related to HAITI HA/DR from 1/13/2010 to 5/26/2010

Report Documentation Page			Form Approved OMB No. 0704-0188		
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE JUN 2012		2. REPORT TYPE		3. DATES COVERED 00-00-2012 to 00-00-2012	
4. TITLE AND SUBTITLE Semantic and Social Networks Comparison for the Haiti Earthquake Relief Operations from APAN Data Sources using Lexical Link Analysis (LLA)			5a. CONTRACT NUMBER		
			5b. GRANT NUMBER		
			5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S)			5d. PROJECT NUMBER		
			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School,Distributed Information Systems Experimentation (DISE) Group,Monterey,CA,93943			8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)			10. SPONSOR/MONITOR'S ACRONYM(S)		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S)		
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES Presented at the 17th International Command & Control Research & Technology Symposium (ICCRTS) held 19-21 June, 2012 in Fairfax, VA.					
14. ABSTRACT The Haiti earthquake relief effort in 2010 involved numerous civil and military organizations engaged in vigorous relief operations, exemplifying rapid deployment of logistics, transport, security and medical supplies. Yet, to determine the effectiveness of these operations, and to understand how these various organizations collaborated, and eventually how to improve the future overall effectiveness of these operations, requires sifting through large volumes of communications and determining a methodology so that analysts may understand how to gain understanding and analyze such data. We have been leveraging ?Lexical Link Analysis? (LLA), as one means to achieve ?System Self-awareness? (SSA), and develop this as a methodology to analyze large text-based data sets, such as the interagency collaborations and problem solving processes using documents obtained from traditional communication platforms as well as new social media. Last year, we reported using LLA to analyze open source data sources of the Haiti relief operations. This year, we will report on the analysis applied to the APAN (The All Partners Access Network) forum discussion data from the relief effort.					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 42	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

- SITREP: ~150 Situation Report documents
- Forum: ~1400 posts from 1/13/2010 to 6/3/2010
- Blogs: ~3900 blog messages

In the previous ICCRTS publication (Zhao et al., 2010), we leveraged “Lexical Link Analysis” (LLA) and a related concept, “System Self-awareness” (SSA) to analyze the open sources of Haiti operations to address the business questions. In this current paper, we seek to leverage the same method and concept but focus on the APAN forum data.

The APAN HA/DR forum data originally included many discussions for the disasters, threats, and risks and was not limited to the Haiti earthquake, we applied a filter to the posts containing a small word list, i.e. Haiti, jacmel, usns, Port-au-Prince, Pap, etc. to exclude the ones with empty message bodies or posted by anonymous users, which resulted in the data set of ~1400 posts. This data set not only revealed the collaborations among military, government, and civil stakeholders in the crisis via social networks, it also recorded the content that reflects how these stakeholders solved problems by addressing issues raised in the thread discussions. A discussion thread, for instance, started when a person raised a question, and other people followed up with answers with different levels of detail.

In this paper, we will discuss two applications of using LLA and the APAN forum data to:

- Discover themes and topics in the discussion forum and sort the importance of the themes (Section 3.2)
- Discover social and semantic networks of organizations that were involved, compare the two networks to obtain insights to answer the following questions:
 - What were the organizations involved in the *important* themes in the Haiti operation (Section 3.3.2 Table 2, 3, 4 and 5)?
 - How do semantic networks suggest more potential collaboration when compared to social networks (Section 3.3.2)?

3 Approaches

3.1 Lexical Link Analysis (LLA)

As in military operations, where the term *situational awareness* is coined, we note that that our efforts can inform *awareness* of analyzed data, in a unique way that helps improve decision maker’s understanding or awareness of its content. We therefore define *awareness* as the cognitive interface between decision makers and a complex system, expressed in a range of terms or “features,” or specific vocabulary or “lexicon,” to describe the attributes and surrounding environment of the system. Specifically, LLA is a form of text mining in which word meanings represented in lexical terms (e.g. word pairs) can be represented as if they are in a community of a word network. Figure 1 demonstrates an example of LLA with connected keywords or concepts extracted from the documents of Maritime Domain Awareness (MDA) technologies. Words are linked as word pairs. Different colors indicate different clusters of word groups. In Figure 2, the word “behavior” is, for example, centered within “suspicious, bad, dangerous, abnormal, usual, and anomalous” and so on, showing the ways to describe “behavior” in this data found from MDA.

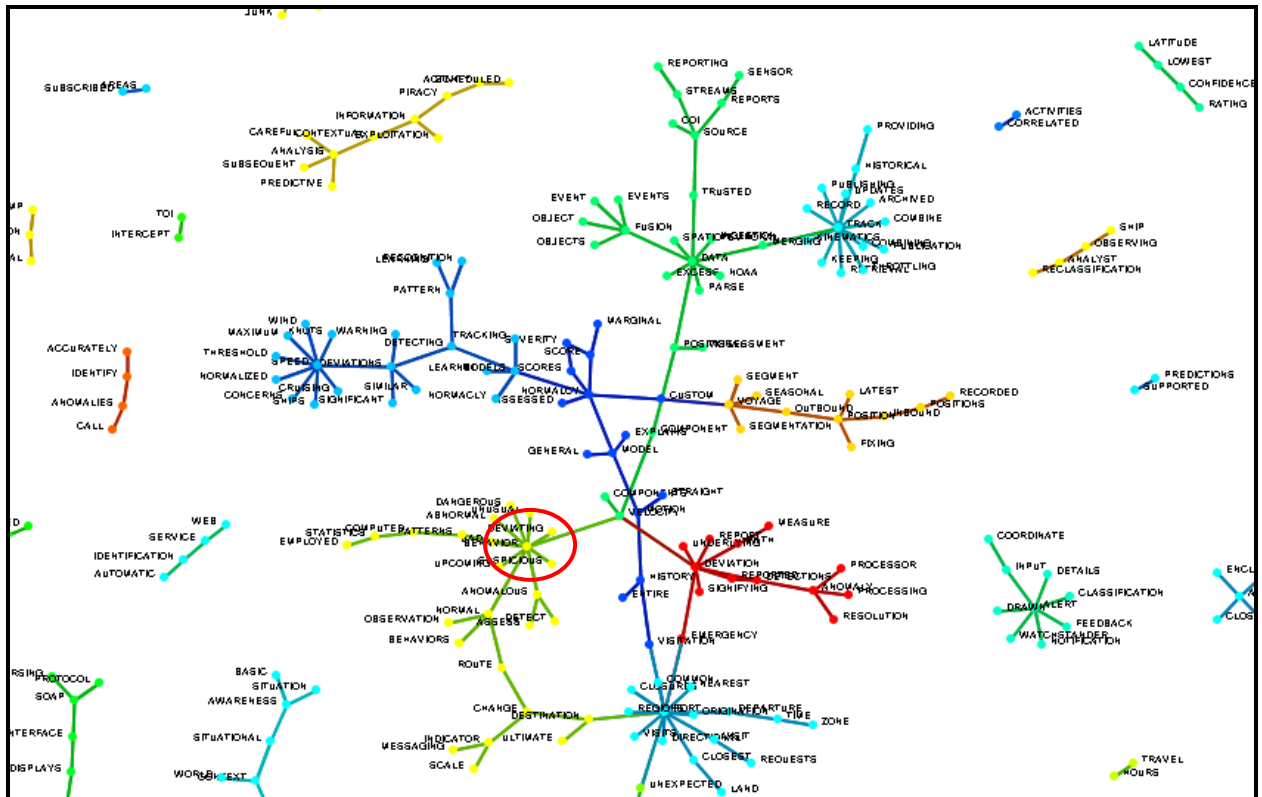


Figure 1: Features Shown as Word Pairs

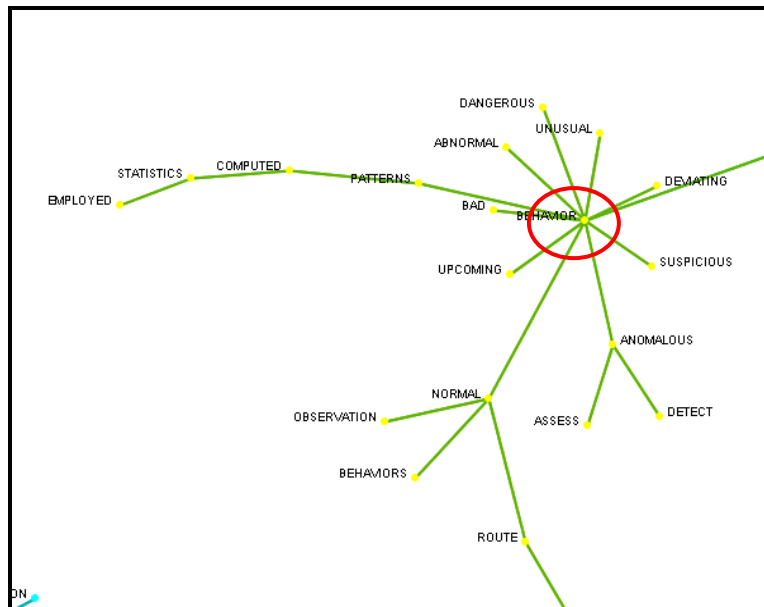


Figure 2: A Single Word Hub Showing Linkage (Detail of Figure 1 in the red circle)

The detailed steps of LLA processing include applying collaborative learning agents (CLA) and generating visualizations, including a lexical network visualization via AutoMap (2009), radar visualization, and matrix visualization (Zhao et al. 2010). The following are the steps for performing an LLA:

- Read each set of documents.
- Select feature-like word pairs.

- Apply a social network community finding algorithm (e.g. Newman grouping method; Girvan et al. 2001) to group the word pairs into themes. A theme includes a collection of lexical word pairs connected each other.
- Compute a “weight” for a theme for the information of a time period, that is, how many word pairs belong to a theme for that time period and for all the time periods.
- Sort theme weights by time, and study the distributions of the themes by time.

3.2 Discover Themes

Figure 3 shows the themes discovered from the APAN forum data set sorted by the number of lexical links (i.e. word pairs) - the column “LLA Counts” in each theme. Figure 4 shows the lexical links in the theme associated with the keywords “WATER, AIR, FOOD.” A theme is summarized generally using a three-word list. Each list of words becomes a word hub with a large number of links (i.e. total degree as a centrality measure). The word hubs are connected at most, a node apart. *Sorted* themes indicate the main and important topics (themes) discussed in the APAN forum. We further divided data between 1/13/2010 to 6/30/2010 into one week as a period after the event until 4/3/2010. The data after 4/3/2010 belongs to one period. The dates are the first days of these weeks.

Table 1 shows the distributions of LLA counts for the themes associated with keywords “WATER, AIR, FOOD” and “PEOPLE, CHILDREN, POST” according to these weeks divided, indicating these activities were continuously performed in many weeks after the event.

Event_Date_Sort	Theme Id	LLA Counts	Theme Keywords
ALL	95	83	WATER,AIR,FOOD
ALL	65	67	PEOPLE,CHILDREN,POST
ALL	6	65	APAN,CONTACT,USER
ALL	people	65	HAITI,OFFICER,PRESIDENT
ALL	64	60	PAP,EMERGENCY,ASSESSMENT,ASSISTANCE
ALL	41	58	INFORMATION,COMMUNITY,REQUESTS
ALL	54	56	MEDICAL,TIME,LIMITED
ALL	29	56	EARTHQUAKE,LEFT,MASSIVE,QUAKE
ALL	46	51	PATIENTS,LONG,CARE
ALL	70	50	RELIEF,FIELD,EFFORT
ALL	73	50	GROUP,SEE,FRENCH
ALL	96	50	DE,POSTED,INITIAL,WORKING
ALL	22	50	COUNTRY,HEALTH,AREAS,CHILE
ALL	place	49	STATE,PORT_AU_PRINCE,CITY
ALL	78	47	BUILDING,HQ,CIVIL,USERS
ALL	86	44	TEAM,COORDINATION,LEAD
ALL	39	42	HOSPITAL,BED,HOUSE
ALL	82	42	SUPPORT,PERSONNEL,CLIENT,UNIT

Figure 3: Sorted Themes for All the Threads

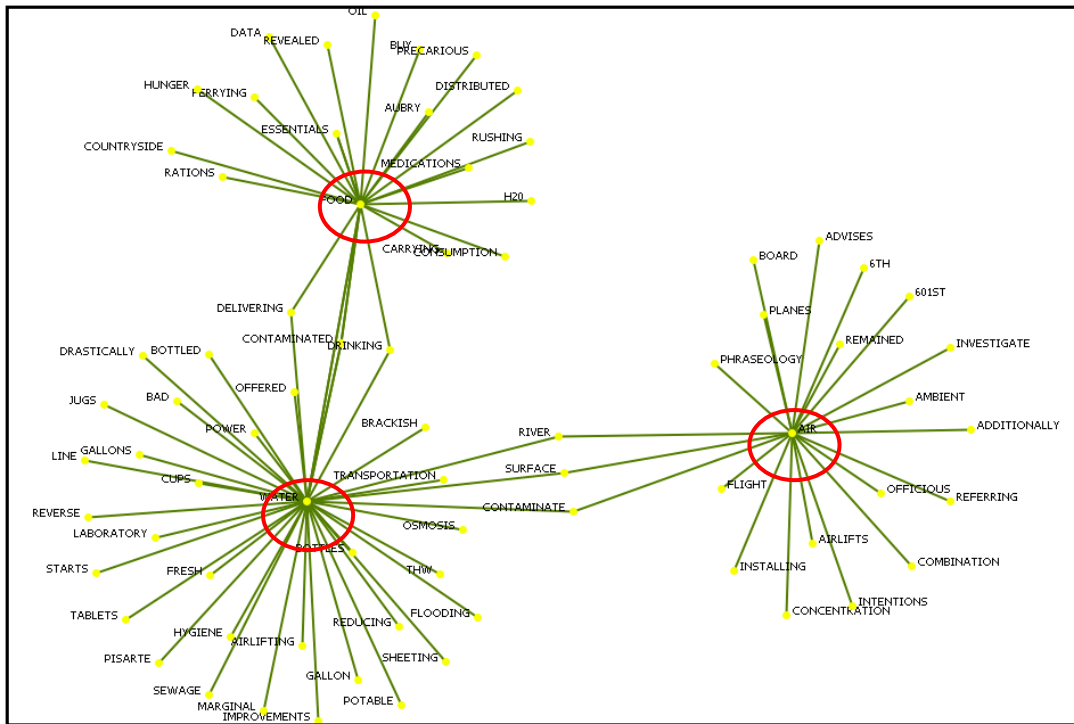


Figure 4: Lexical links for the theme associated with keywords “WATER, AIR, FOOD”.

Table 1: Distributions of LLA Counts in the Weeks

Dates	Theme ID	LLA Counts	Theme Keywords
ALL	95	83	WATER,AIR,FOOD
1/20/2010	95	39	WATER,AIR,FOOD
1/13/2010	95	37	WATER,AIR,FOOD
1/27/2010	95	20	WATER,AIR,FOOD
3/10/2010	95	7	WATER,AIR,FOOD
2/10/2010	95	3	WATER,AIR,FOOD
2/17/2010	95	3	WATER,AIR,FOOD
2/24/2010	95	3	WATER,AIR,FOOD
2/3/2010	95	2	WATER,AIR,FOOD
3/3/2010	95	2	WATER,AIR,FOOD
ALL	65	67	PEOPLE,CHILDREN,POST
1/20/2010	65	35	PEOPLE,CHILDREN,POST
1/13/2010	65	19	PEOPLE,CHILDREN,POST
1/27/2010	65	8	PEOPLE,CHILDREN,POST
2/10/2010	65	7	PEOPLE,CHILDREN,POST
3/3/2010	65	4	PEOPLE,CHILDREN,POST
2/24/2010	65	3	PEOPLE,CHILDREN,POST
2/3/2010	65	2	PEOPLE,CHILDREN,POST
2/17/2010	65	1	PEOPLE,CHILDREN,POST
3/10/2010	65	1	PEOPLE,CHILDREN,POST
3/21/2010	65	1	PEOPLE,CHILDREN,POST

The themes found using LLA are consistent with the existing doctrines of the roles of military and non-government organizations (NGO) might engage in if a disaster like the Haiti earthquake occurs. For example, United States Agency for International Development (USAID) specifies a *cluster system* for what various organizations should perform, where military is employed in the tasks of handling logistics and relief supplies (USAID, 2010).

When looking at the detail and the themes produced using LLA, the actual focuses of the APAN forum were revealed. Figure 5 shows the themes sorted using the numbers of lexical links in the first and second weeks of the Haiti operation (1/13/2010 to 1/19/2010 and 1/20/2010 to 1/26/2010). Compared to the overall themes in Figure 3, themes on the first two weeks were different, indicating priorities in the initial phase of the relief effort were different than in the total relief effort. Figure 7 and Figure 6 show the lexical links for the themes associated with the keywords “HIGH, RATES”, “AIRCRAFT, GROUND, ASSIST, ORDERED, USNS_COMFORT”, and “SUPPORT, PERSONNEL, CLIENT, UNIT” are not found among the top ten themes in Figure 3. These were examples of the themes that were unique to the first week. For example, the theme “HIGH, RATES” was related to the fact that the US government provided the analysis and information about high fatalities, high risks and high disease rates that might be expected after the event. This suggests these important issues were heavily discussed and collaborated in the time of frame of two weeks. The focuses of the APAN forum activities have been shifted from the logistics centric first week to medical and patient centric second week in the context of the overall efforts.

Event_Date_Sort	Theme Id	LLA Counts	Theme Keywords
1/13/2010	ALL	630	
1/13/2010	95	37	WATER,AIR,FOOD
1/13/2010	37	25	HIGH,RATES
1/13/2010	5	22	AIRCRAFT,GROUND,ASSIST,ORDERED,USNS_COMFORT
1/13/2010	6	21	APAN,CONTACT,USER
1/13/2010	82	21	SUPPORT,PERSONNEL,CLIENT,UNIT
1/13/2010	65	19	PEOPLE,CHILDREN,POST
1/13/2010	people	18	HAITI,OFFICER,PRESIDENT
1/13/2010	64	18	PAP,EMERGENCY,ASSESSMENT,ASSISTANCE
1/13/2010	79	18	STAFF,VERSION,BATTLE,LOGS,RECORDS
1/13/2010	69	18	PROVIDE,DEAD,NEEDED
1/13/2010	39	17	HOSPITAL,BED,HOUSE

Event_Date_Sort	Theme Id	LLA Counts	Theme Keywords
1/20/2010	95	39	WATER,AIR,FOOD
1/20/2010	65	35	PEOPLE,CHILDREN,POST
1/20/2010	people	33	HAITI,OFFICER,PRESIDENT
1/20/2010	46	32	PATIENTS,LONG,CARE
1/20/2010	29	26	EARTHQUAKE,LEFT,MASSIVE,QUAKE
1/20/2010	54	25	MEDICAL,TIME,LIMITED
1/20/2010	41	24	INFORMATION,COMMUNITY,REQUESTS
1/20/2010	6	23	APAN,CONTACT,USER
1/20/2010	70	23	RELIEF,FIELD,EFFORT
1/20/2010	72	23	RESPONSE,HELICOPTER,PART

Figure 5: Themes sorted for the first week starting on 1/13/2010 (to 1/19/2010) and the second week starting on 1/20/2010 (to 1/26/2010).

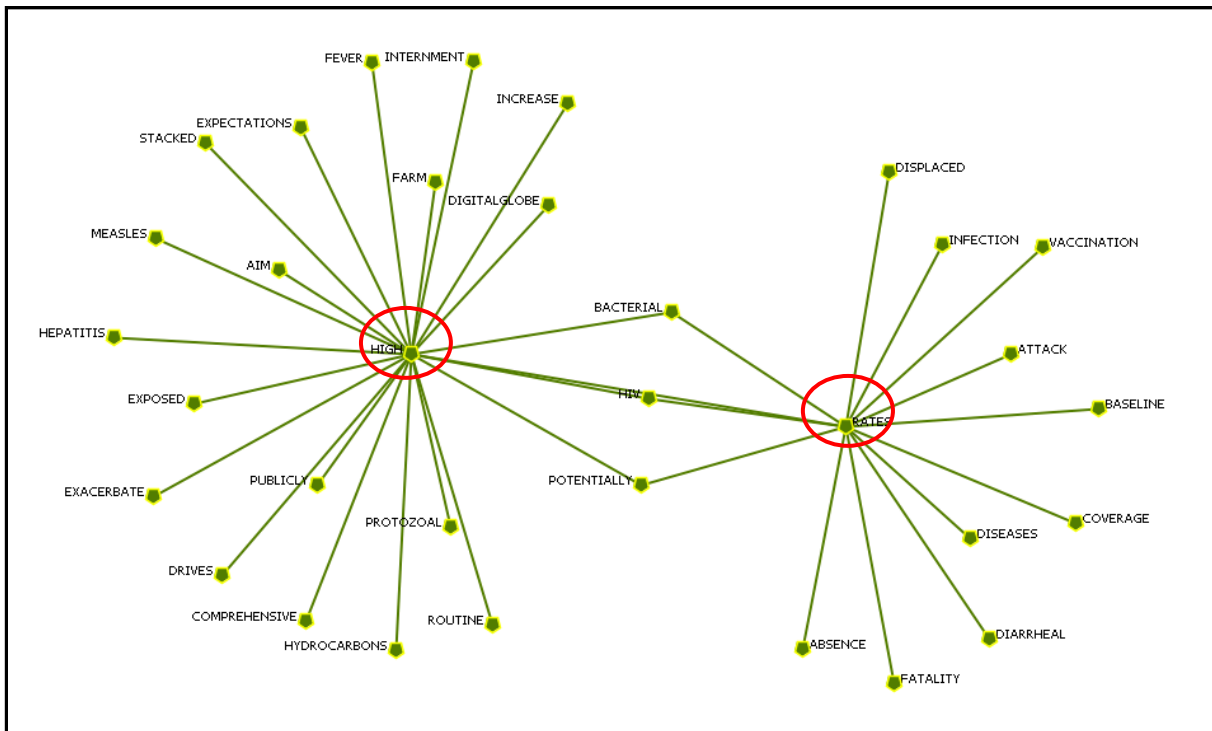


Figure 6: Lexical links for the theme associated with keywords “HIGH, RATES”.

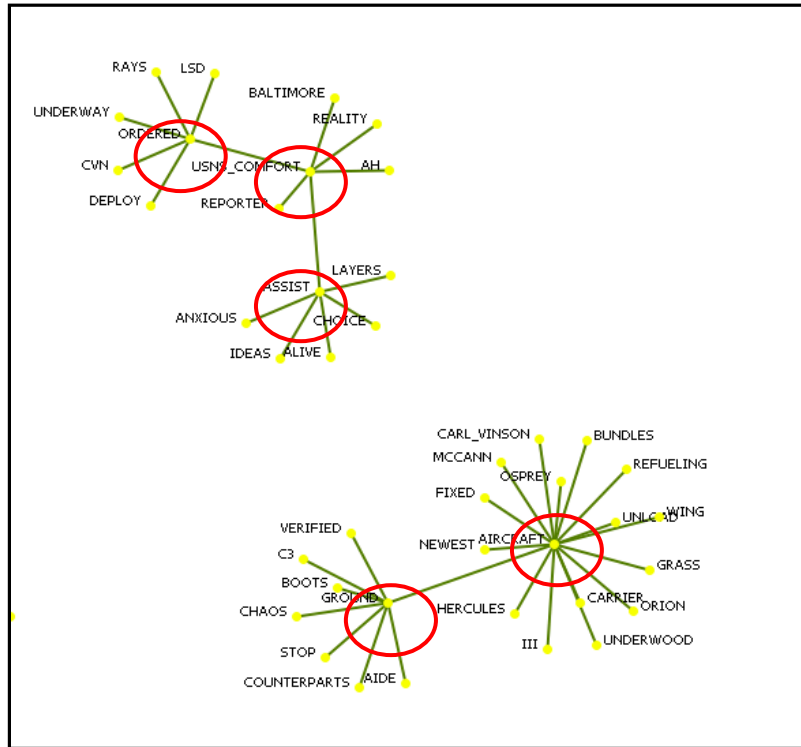


Figure 7: Lexical links for the theme associated with keywords “AIRCRAFT, GROUND, ASSIST, ORDERED, USNS_COMFORT”.

Figure 8 shows the overlapping of top ten themes sorted by the LLA counts overall (ALL) and the first (Week 1) and second weeks (Week 2). Our observations reveal that:

- Six out of ten top themes (60%) in the top themes of all are also in the top themes of the first weeks. Likewise, seven of ten top themes in the first two weeks are also in the top overall themes, indicating the LLA counts are indeed valid measures to sort out the main and important topics and themes. In this example, the important topics were handled in the first two weeks of the event.
- Some of the important topics or themes which are picked out by the LLA counts are not shown in the first two weeks. For example, “MEDICAL, TIME, LIMITED” and “PAP, EMERGENCY, ASSESSMENT, ASSISTANCE” may not be feasible to address as the high priority items for the first two weeks. While the themes “INFORMATION, COMMUNITY, REQUESTS” and “RELIEF, FIELD, EFFORT” can be themes that are important across all the time frames.

	Theme Keywords		Theme Keywords
All	WATER,AIR,FOOD	Week 1	WATER,AIR,FOOD
	PEOPLE,CHILDREN,POST		HIGH,RATES
	APAN,CONTACT,USER		AIRCRAFT,GROUND,ASSIST,ORDERED,USNS_COMFORT
	HAITI,OFFICER,PRESIDENT		APAN,CONTACT,USER
	PAP,EMERGENCY,ASSESSMENT,ASSISTANCE		SUPPORT,PERSONNEL,CLIENT,UNIT
	INFORMATION,COMMUNITY,REQUESTS	Week 2	WATER,AIR,FOOD
	MEDICAL,TIME,LIMITED		PEOPLE,CHILDREN,POST
	EARTHQUAKE,LEFT,MASSIVE,QUAKE		HAITI,OFFICER,PRESIDENT
	PATIENTS,LONG,CARE		PATIENTS,LONG,CARE
	RELIEF,FIELD,EFFORT		EARTHQUAKE,LEFT,MASSIVE,QUAKE

Figure 8: The overlapping of top ten themes sorted by the LLA counts overall (ALL) and the first and second weeks.

Much of text analysis depends on initially searching the available internet. At this point, our efforts are sometimes compared to those of a typical search engine. One of the disadvantages of conventional search engines is that they typically sort documents based on the popularity of documents among linked documents, not based on semantics. Therefore, it does not satisfy complete search needs nor determine relevance if the links among the documents are not available. For example, the content in the forum is not cross-linked, therefore, the discovered or *revealed* topics or themes cannot be found as prioritized results, if conventional search engines are used.

The LLA approach is more properly related to Latent Semantic Analysis (LSA; (Dumais, Furnas, Landauer, Deerwester, & Harshman, 1988; Gorman, 2003; Letsche, 1997) and Probabilistic Latent Semantic Analysis (PLSA). In the LSA approach, a term-document matrix is the starting point for analysis. The elements of the term-document or feature-object (term as feature and document as object) matrix are the occurrences of each word in a particular document, *i.e.* $A = [a_{ij}]$, where a_{ij} denotes the frequency in which term j occurs in document i . The term-document matrix is usually sparse. LSA uses singular value decomposition (SVD) to reduce the dimensionality of the term-document matrix. SVD cannot be applied to the cases where the vocabulary (the unique number of terms) in the document collection is large. LSA has been widely used to improve information indexing, search/retrieval and text categorization.

A recent development related to this method is called Latent *Dirichlet* allocation (LDA; Blei, Ng, & Jordan, 2003), which is a generative probabilistic model of a corpus. In LDA, a document is considered to be composed of a collection of words—a “bag of words,” where word order and grammar are not considered important. The basic idea is that documents are represented as random mixtures over latent topics, where each topic is characterized by a statistical distribution (Dirichlet distribution) over the corpus. Our theme generation from LLA is different than LDA, in which a collection of lexical terms are connected to each other semantically, as if they are in a social community, and social network grouping methods are used to group the words. Our method is easily scaled to analyze a large vocabulary and is generalizable to any sequential data.

3.3 *Semantic and Social Networks*

3.3.1 *Extracting Social Networks*

Current research of social network analysis mostly focuses on using a simple, weighted, or arrowed links between two objects, such as people or organizations, regardless of the contents (Wasserman, 1994). The so-called study of centrality (Girvan 2002) has been a focal point for the social network structure studies in recent years. Finding the *centrality* of a network lends insight into the various roles and groupings, such as the connectors (for example, mavens, leaders, bridges, isolated nodes), the clusters (and who is in them), the network core, and its periphery.

We extract the social networks based how an organization was engaged in a discussion of an issue, *i.e.*, a discussion thread. For example, in Figure 9, Organization “DoD Air Force,” “SSSG,” etc. were engaged in the discussion in Thread 5741, regarding the subject “mtpg airfield sitrep”, therefore, we assume these organizations were working together and formed social network connections.

Figure 10 shows an organization and thread matrix which lists how many times an organization was involved in the thread discussion. This matrix was then used to generate the social networks of participating organizations - one organization was connected to another if they were both involved in at least one of the same thread discussions.

With the themes separated from the theme discovery process, we were able to look at the organization connections within each theme. While traditional social networks typically look at a social network as a whole, Figure 11 shows a social network of the organizations involved in the Haiti operation only for the

theme “WATER, AIR, FOOD”. The light blue connections represent the social network connections that are overlapped with the semantic networks, indicating that these organizations not only appeared together in the same threads, but also discussed concepts and contents similar to each other which were the basis to form semantic connections (see Section 3.3.2.). Figure 12 shows the organizations illustrated in a node cloud, where the ones with larger fonts are the organizations with higher centralities measured in *total degrees* in Figure 11.

1	Thread	Level	Time	User	Organization	Content
43	5722	2	1/14/2010	2205	DoD U S Southern	re: attention!!! notice to airmen notam " That is correct. In order to gain aerial access
44	5722	2	1/15/2010	3342	Department of	re: attention!!! notice to airmen notam " I suggest that MTPP be used as a tag for all
45	5729	1	1/14/2010	3645	DoD Air Force	pap ground support I noticed this morning that there was a 60K main deck loader at
46	5736	1	1/14/2010	2205	DoD U S Southern	port-au-prince harbor port damage photos " [] [] [] [] [] "
47	5737	1	1/14/2010	2205	DoD U S Southern	rft on behalf of mlgp haiti " Is there any info on a USCG helo flight inbound from Santo
48	5737	2	1/14/2010	2112	DoD US Southern	rft on behalf of mlgp haiti " Heather, dont know, but we should send invite using the
49	5737	3	1/14/2010	2205	DoD U S Southern	rft on behalf of mlgp haiti Yes! Lets do it. can you get them the directions or do you
50	5737	4	1/15/2010	2112	DoD US Southern	rft on behalf of mlgp haiti " Heather, I dont want to step out of my lane. I think it
51	5741	1	1/14/2010	3645	DoD Air Force	mtpp airfield sitrep Can anyone provide any information regarding the Airfield
52	5741	2	1/15/2010	3342	Department of	mtpp airfield sitrep " I suggest that MTPP be used as a tag for all NOTAMs, blog posts,
53	5741	2	1/14/2010	3545	SSSG	mtpp airfield sitrep " Southern Command Press Release, QUOTE: "airfield has power
54	5741	3	1/15/2010	4038	Telecommunicatio	mtpp airfield sitrep " Both should be used, since Civilian Operators use NOTAMs as
55	5741	4	1/15/2010	3342	Department of	mtpp airfield sitrep Good point.
56	5743	1	1/14/2010	2406	Integrated ICT	usar activity as of 142130z " *summary by* veronika wolf ocha - 14-jan-2010 21:53
57	5744	1	1/14/2010	2406	Integrated ICT	logistical activity as of 142000est " *summary by* veronika wolf ocha - 14-jan-2010
58	5745	1	1/14/2010	2406	Integrated ICT	coordination activity as of 141641z " *summary by* veronika wolf ocha - 14-jan-2010
59	5746	1	1/14/2010	2406	Integrated ICT	situation activity as of 150137z " re msf facilities haiti: msf treats more than 1,000

Figure 9: An APAN forum discussion thread example

1		10022	10023	10033	10100	10113	10393	10406	10432	10473	10484	10491
78	MilitaryUSANORTHCOM_	0	0	0	0	0	0	0	0	0	0	0
79	MilitaryUSANORTHCOM_Joint	0	0	0	0	0	0	0	0	0	0	0
80	MilitaryUSANationalDefenseUniversity_	0	0	0	0	0	0	0	0	0	0	0
81	MilitaryUSANavy_	0	0	0	0	0	0	0	0	0	0	0
82	MilitaryUSANavy_Command	0	0	0	0	0	0	0	0	0	0	0
83	MilitaryUSANavy_InformationManagement	0	0	0	0	0	0	0	0	0	0	0
84	MilitaryUSANavy_Infrastructure	0	0	0	0	0	0	0	0	0	0	0
85	MilitaryUSANavy_Intelligence	0	0	0	0	0	0	0	0	0	0	0
86	MilitaryUSANavy_Joint	0	0	0	0	0	0	0	0	0	0	0
87	MilitaryUSANavy_JointIT	0	0	0	0	0	0	0	0	0	0	0
88	MilitaryUSANavy_LanguageInstitute	0	0	0	0	0	0	0	0	0	0	0
89	MilitaryUSANavy_Planning	0	0	0	0	0	0	0	0	0	0	0
90	MilitaryUSAPACOM_	0	0	0	0	0	0	0	0	0	0	0
91	MilitaryUSAPACOM_APAN	0	0	0	1	0	0	0	0	0	0	0
92	MilitaryUSAPACOM_InformationManagement	1	0	0	0	0	0	0	0	0	0	0
93	MilitaryUSAPACOM_Joint	0	0	0	0	0	0	0	0	0	0	0
94	MilitaryUSAPACOM_Planning	0	0	0	0	0	0	0	0	0	0	0
95	MilitaryUSAPACOM_Simulation	0	0	0	2	0	0	0	0	0	0	2
96	MilitaryUSASOUTHCOM_	0	0	0	0	0	0	1	1	0	0	0
97	MilitaryUSASOUTHCOM_Command	0	0	0	0	0	0	0	0	0	0	0
98	MilitaryUSASOUTHCOM_GIS	0	0	0	0	0	0	0	0	0	0	0
99	MilitaryUSASOUTHCOM_InformationManagement	0	0	0	0	0	0	0	0	0	0	0
100	MilitaryUSASOUTHCOM_Intelligence	0	0	0	0	0	0	0	0	0	0	0
101	MilitaryUSASOUTHCOM_Joint	0	0	0	0	0	0	0	0	0	0	0
102	MilitaryUSASOUTHCOM_ManagementSOUTHCOM	0	0	0	0	0	0	0	0	0	0	0
103	MilitaryUSASOUTHCOM_Partnering	0	0	0	0	0	0	0	0	0	0	0
104	MilitaryUSASOUTHCOM_Planning	0	0	0	0	0	0	0	0	0	0	0
105	MilitaryUSASOUTHCOM_Security	0	0	0	0	0	0	0	0	0	0	0
106	MilitaryUSASOUTHCOM_SpecialOperations	0	0	0	0	0	0	0	0	0	0	0

Figure 10: Organization and thread matrix

3.3.2 Extracting Semantic Networks

A semantic network is not necessarily a real connection between two organizations in real-life, instead, it is a network measuring how much similarity of content that two organizations present in the forum, a measure of *like-mindedness*. It is reasonable to assume that like-minded organizations do use *many* of the same terminologies in their communications. For that, *like-mindedness* is closer in meaning to *like-wordedness*, which is the reason to cluster them together. In this case, by using LLA, a semantic network for two organizations can be generated by counting the common word hubs in each theme for the contents of two organizations. In Figure 13, Organization “PrivateUSACopernicusEnergyInc” shares the word hubs with various other organizations such as “MilitaryUSAArmy” and “MilitarySouthAfricaArmyAZ”. Word hubs in common are “Information--”, “Small--” and “Power--”, etc. The words such as “_Energy” and “_Engineering” attached next to the organizations are the self-claimed function areas, possibly reflecting their motivations to join the discussions. The weights next to the word hubs, such as “Information--”, “Small--” and “Power--”, are the numbers of lexical links in the themes to which the word hubs belong. They indicate how important these word hubs, or semantic connections are for the two organizations, as they are sorted in Figure 13. In this example, the word hub “Power--” has the highest weight 302.0. We also submit that some like-minded organizations might use different terminologies in their activities. For example, the military and emergency response organizations often use different words, yet for the same meaning and organizations. The military especially seems to use many acronyms instead of words. In this case, LLA often places the different terminologies in the same theme although they do not share the same word hubs or appear in disconnected components of word networks such as “CVN” and “Carl_Vinson” shown in Figure 7. In this case, the similarity of two organizations can be counted by the words in the same themes. This paper only addresses the like-mindedness in the former case using similarities of word hubs.

```
PrivateUSACopernicusEnergyInc_Energy MilitaryUSAArmy_Engineering 5 ; INFORMATION--(189.00)
PrivateUSACopernicusEnergyInc_Energy MilitarySouthAfricaArmyAZ_ 5 ; INFORMATION--(189.00)
PrivateUSACopernicusEnergyInc_Energy MilitaryUSASOUTHCOM_ 5 ; INFORMATION--(189.00)
PrivateUSACopernicusEnergyInc_Energy MilitaryCanadaNationalDefenseCA_ 5 ; INFORMATION--(189.00)
PrivateUSACopernicusEnergyInc_Energy MilitaryUSAPACOM_Simulation 5 ; INFORMATION--COMMUNITY(189.00)
PrivateUSACopernicusEnergyInc_Energy GovernmentUSADOD_ 5 ; SMALL--(208.00)
PrivateUSACopernicusEnergyInc_Energy MilitaryUSASOUTHCOM_ManagementSOUTHCOM 5 ; INFORMATION--(189.00)
PrivateUSACopernicusEnergyInc_Energy PrivateUSARapidResponseConsulting_TelecommunicationsDisasterResponse 5 ; INFORMATION--(189.00)
PrivateUSACopernicusEnergyInc_Energy MilitaryUSANavy_InformationManagement 5 ; POWER--(302.00)
PrivateUSACopernicusEnergyInc_Energy NGOSAWorldCares_DisasterResponseNetworks 5 ; INFORMATION--(189.00)
PrivateUSACopernicusEnergyInc_Energy MilitaryUSADOD_SimulationJoint 5 ; SMALL--(208.00)
PrivateUSACopernicusEnergyInc_Energy MilitaryUSAPACOM_APAN 5 ; INFORMATION--REQUESTS(189.00)
PrivateUSACopernicusEnergyInc_Energy MilitaryUSADOD_IntelligenceJoint 5 ; INFORMATION--(189.00)
PrivateUSACopernicusEnergyInc_Energy PrivateUSAPaxModel_Reconstruction 5 ; INFORMATION--CONFIDENTIAL(189.00)
MilitaryUSADOD_ITJoint MilitaryUSADOD_InformationSharing 5 ; SUSAN--(98.00)
MilitaryUSADOD_ITJoint MilitaryUSAEuropeanCommand_ 5 ; SUSAN--(68.00)
PrivateUSAemicus.com_DisasterResponse MilitaryUSASOUTHCOM_Security 5 ; BOBBI--(1.00); JACMEL--(111.00)
PrivateUSAemicus.com_DisasterResponse MilitaryUSAArmy_Engineering 5 ; BACKGROUND--(189.00)
PrivateUSAemicus.com_DisasterResponse NGOSAPROJECTKIDInc_Adoption 5 ; FOOD--(302.00); FOOD--WATER(302.00); JACMEL--(111.00)
PrivateUSAemicus.com_DisasterResponse MilitarySouthAfricaArmyAZ_ 5 ; FOOD--(302.00); REQUEST--(102.00)
PrivateUSAemicus.com_DisasterResponse MilitaryUSASOUTHCOM_ 5
GEAR--(97.00); FOOD--(302.00); GOOGLE--(225.00); REQUEST--(102.00); REQUEST--EXPLANATION(102.00); TRANSPORTING--(111.00); BACKGROUND--(189.00)
PrivateUSAemicus.com_DisasterResponse MilitaryUSAPACOM_Simulation 5 ; FOOD--(302.00)
PrivateUSAemicus.com_DisasterResponse GovernmentUSADOD_ 5
FOOD--DRINKING, CONTAMINATED(302.00); FOOD--WATER(302.00); FOOD--WATER, DRINKING, CONTAMINATED(302.00); FOOD--(302.00)
```

Figure 13: Using LLA to form semantic network connections

Figure 14 shows the social networks (blue) in Figure 11 overlaid with the semantic networks (green) which are associated with the theme with the keywords “WATER, AIR, FOOD.” The light blue connections are the ones overlapped in the two networks analyzed.

Note that in Figure 14, a few nodes with only semantic connections, i.e. only green links can be found: These are the organizations which had no social connections with other organizations in the theme, however, they had discussion content which are similar to the ones of other organizations, therefore, they are the potential

high-value targets to be included in the collaborations. Figure 15 shows the word hubs which are the basis for one of the semantic network nodes: PrivateUSAstrategos_SecurityContracting.

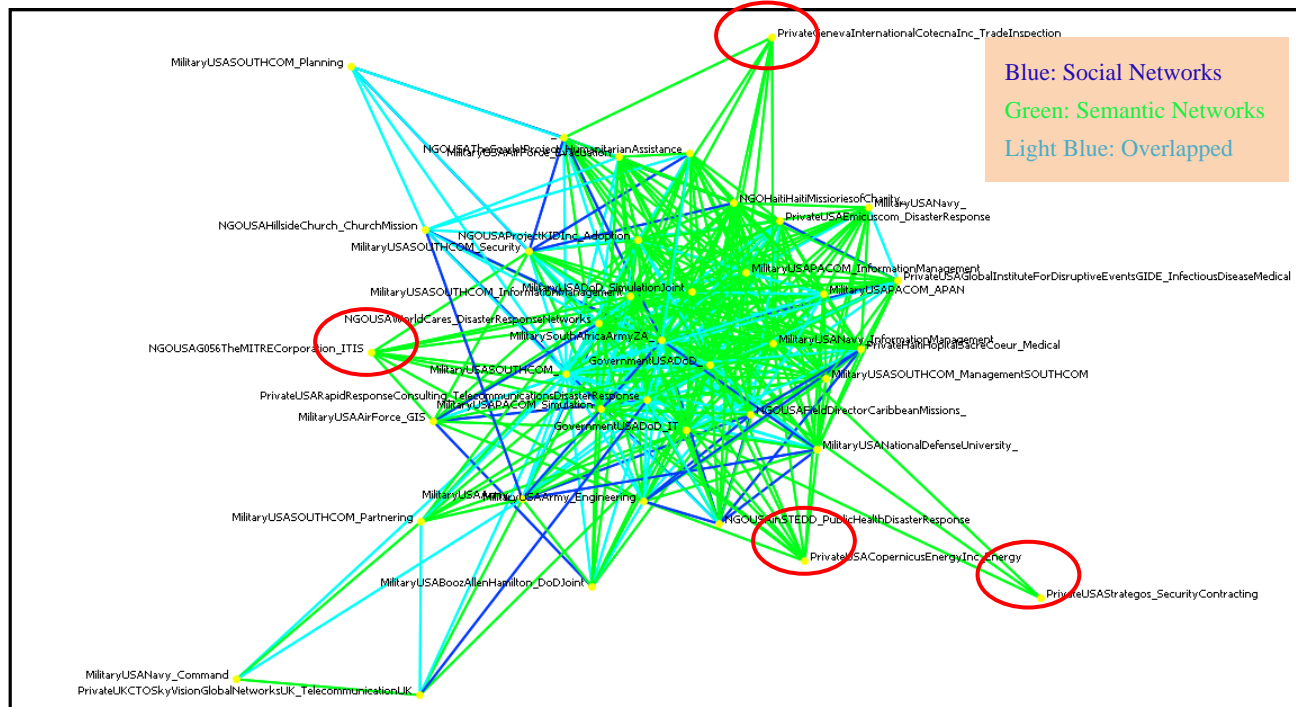


Figure 14: Semantic networks (green) overlaid on the top of the social networks for the theme associated with keywords “WATER, AIR, FOOD”.

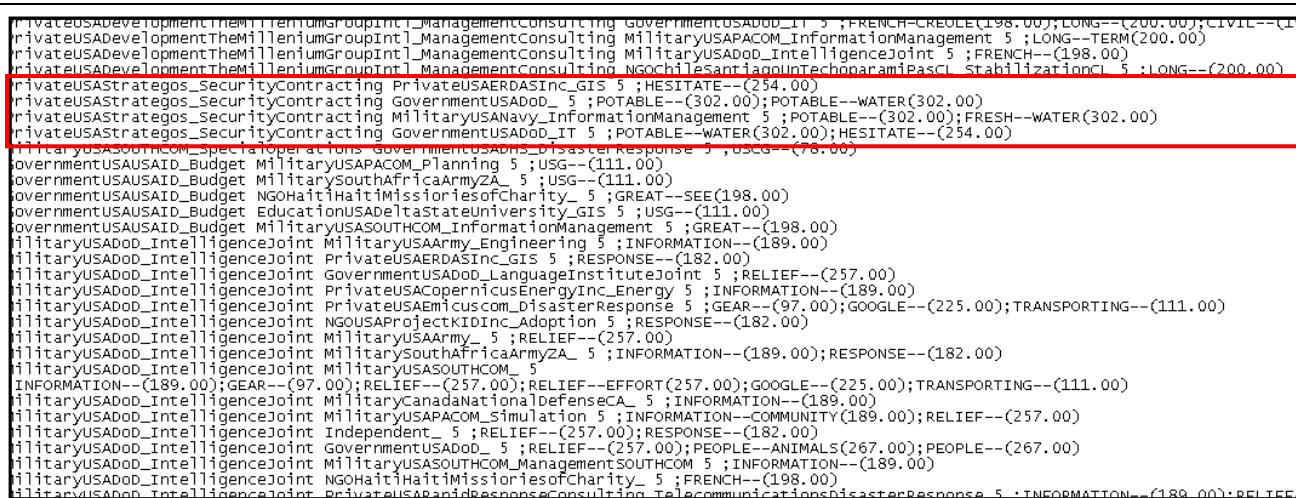


Figure 15: The word hubs for PrivateUSAstrategos_SecurityContracting in the semantic network.

Table 2 shows the Organizations involved in the theme “WATER, AIR, FOOD”. The same analyses can be done for other themes. Table 3, Table 4 and Table 5 show the organizations involved in the three other important themes “HIGH, RATES”, “AIRCRAFT, GROUND, ASSIST, ORDERED, USNS_COMFORT”, and “MEDICAL, TIME, LIMITED”. The highlighted ones are the ones suggested by semantic networks, which have no social network connections with other organizations.

Table 2: Organizations involved in the theme “WATER, AIR, FOOD”

GovernmentUSADoD_IT
MilitarySouthAfricaArmyZA_
MilitaryUSAAirForce_Evacuation
MilitaryUSAAirForce_GIS
MilitaryUSAArmy_
MilitaryUSAArmy_Engineering
MilitaryUSABoozAllenHamilton_DoDJoint
MilitaryUSADoD_SimulationJoint
MilitaryUSANationalDefenseUniversity_
MilitaryUSANavy_
MilitaryUSANavy_Command
MilitaryUSANavy_InformationManagement
MilitaryUSAPACOM_APAN
MilitaryUSAPACOM_InformationManagement
MilitaryUSAPACOM_Simulation
MilitaryUSASOUTHCOM_
MilitaryUSASOUTHCOM_InformationManagement
MilitaryUSASOUTHCOM_ManagementSOUTHCOM
MilitaryUSASOUTHCOM_Partnering
MilitaryUSASOUTHCOM_Planning
MilitaryUSASOUTHCOM_Security
NGOHaitiHaitiMissionariesofCharity_
NGOUSAFieldDirectorCaribbeanMissions_
NGOUSAG056TheMITRECorporation_ITIS
NGOUSAHillsideChurch_ChurchMission
NGOUSAinSTEDD_PublicHealthDisasterResponse
NGOUSAProjectKIDInc_Adoption
NGOUSATheScarletProject_HumanitarianAssistance
NGOUSAWorldCares_DisasterResponseNetworks
PrivateGenevaInternationalCotecnaInc_TradeInspection
PrivateHaitiHopitalSacreCoeur_Medical
PrivateUKCTOSkyVisionGlobalNetworksUK_TelecommunicationUK
PrivateUSACopernicusEnergyInc_Energy
PrivateUSAEmicuscom_DisasterResponse
PrivateUSAGlobalInstituteForDisruptiveEventsGIDE_InfectiousDiseaseMedical
PrivateUSARapidResponseConsulting_TelecommunicationsDisasterResponse
PrivateUSAstrategos_SecurityContracting

Table 3: Organizations involved in the theme “HIGH, RATES”

GovernmentUSADoD_
GovernmentUSADoD_IT
MilitaryUSANationalDefenseUniversity_
MilitaryUSAPACOM_InformationManagement
MilitaryUSAPACOM_Simulation
MilitaryUSASOUTHCOM_
PrivateUSACopernicusEnergyInc_Energy
PrivateUSADigitalGlobeInc_GISInternationalDefenseSales
PrivateUSAERDASInc_GIS

Table 4: Organizations involved in the theme “AIRCRAFT, GROUND, ASSIST, ORDERED, USNS_COMFORT

EducationUSAUniversityofHawaii_DisasterResponse
GovernmentUSADoD_IT
MilitarySouthAfricaArmyZA_
MilitaryUSAAirForce_Evacuation
MilitaryUSAAirForce_GIS
MilitaryUSAAirForce_ManagementAirforce
MilitaryUSABoozAllenHamilton_DoDJoint
MilitaryUSADoD_GIS
MilitaryUSAPACOM_APAN
MilitaryUSAPACOM_InformationManagement
MilitaryUSASOUTHCOM_
MilitaryUSASOUTHCOM_InformationManagement
MilitaryUSASOUTHCOM_Planning
NGOUSAHillsideChurch_ChurchMission
NGOUSAProjectKIDInc_Adoption
NGOUSAWorldCares_DisasterResponseNetworks
PrivateUSAASIGroup_Security
PrivateUSAERDASInc_GIS
PrivateUSAGlobalInstituteForDisruptiveEventsGIDE_InfectiousDiseaseMedical
PrivateUSARapidResponseConsulting_TelecommunicationsDisasterResponse

Table 5: Organizations involved in the theme “MEDICAL, TIME, LIMITED”

GovernmentUSADoD_
GovernmentUSADoD_IT
GovernmentUSADoD_JointAirlift
GovernmentUSADoD_JointCivilInformationManagement
MilitaryUSAAirForce_GIS
MilitaryUSAArmy_MedicineSurgeonGeneral
MilitaryUSADoD_InformationManagementJoint
MilitaryUSADoD_SimulationJoint
MilitaryUSANationalDefenseUniversity_
MilitaryUSANavy_Joint
MilitaryUSAPACOM_APAN
MilitaryUSAPACOM_InformationManagement
MilitaryUSAPACOM_Simulation
MilitaryUSASOUTHCOM_
MilitaryUSASOUTHCOM_InformationManagement
MilitaryUSASOUTHCOM_Intelligence
MilitaryUSASOUTHCOM_Security
NGOHaitiHaitiMissionariesofCharity_
NGOUSAHillsideChurch_ChurchMission
NGOUSAPartnersinHealth_Logistics
NGOUSAProjectKIDInc_Adoption
NGOUSAWorldCares_DisasterResponseNetworks
PrivateHaitiHopitalSacreCoeur_Medical
PrivateUSAGlobalHealthInitiativesInc_Medicine
PrivateUSARapidResponseConsulting_TelecommunicationsDisasterResponse

The core innovation here is to use the LLA method to analyze the forum content, therefore generate alternative networks, i.e. semantic networks that can be used to compare with the social networks. When comparing semantic networks with the social networks, we showed how to measure synergies among organizations and how the synergies can be varied among different themes. We also showed the potential to rank high-value organizations earlier in the collaboration process based on their common interests from the semantic network analysis for the discovered themes.

4 Additional Observations

As our goal is to use LLA to understand the synergy, efficiency, and competency of an interagency operation that involves many organizations, we use lexical links as metrics to measure synergy-like values. One of such synergy is measured using the ratio of the links in the overlapped parts of social networks and semantic networks.

Figure 16 shows a holistic view of social networks and semantic networks for the organizations associated with the theme “WATER, AIR, FOOD”, where the red connections linking the organizations that have content related to the theme. In Figure 17, the table shows the summary of connection counts for the top ten themes found in the overall forum and on the first two weeks (highlighted in yellow). The table is sorted according to the Synergy column, computed as the ratio of from the columns “Social (Blue and Light Blue)” and “Overlapped (Light Blue)”. It also lists the number of potential semantic links (green) and number of total organizations (red) involved in the theme. Our observations are summarized as follows:

1. There are fewer number of organizations involved on the two weeks than in the overall forum, where the synergies seem go above the average (>80%) for some specific and time-critical themes, for example, “HIGH,RATES”, “PEOPLE,CHILDREN,POST”, “APAN,CONTACT,USER”, “AIRCRAFT,GROUND,ASSIST,ORDERED,USNS_COMFORT”, “SUPPORT,PERSONNEL,CLIENT,UNIT”, and “EARTHQUAKE,LEFT,MASSIVE,QUAKE”
2. There are some themes with relatively low synergies overall and on the two weeks, for example, “WATER, AIR, FOOD”, “PATIENTS, LONG, CARE”, “HAITI, OFFICER, PRESIDENT”. This may indicate lack of knowledge or expertise in these areas in general.
3. Comparing the two weeks and the average, the synergies among organizations seem correlated with the number of organizations involved in the collaborations: the themes with fewer organizations involved have higher synergies.

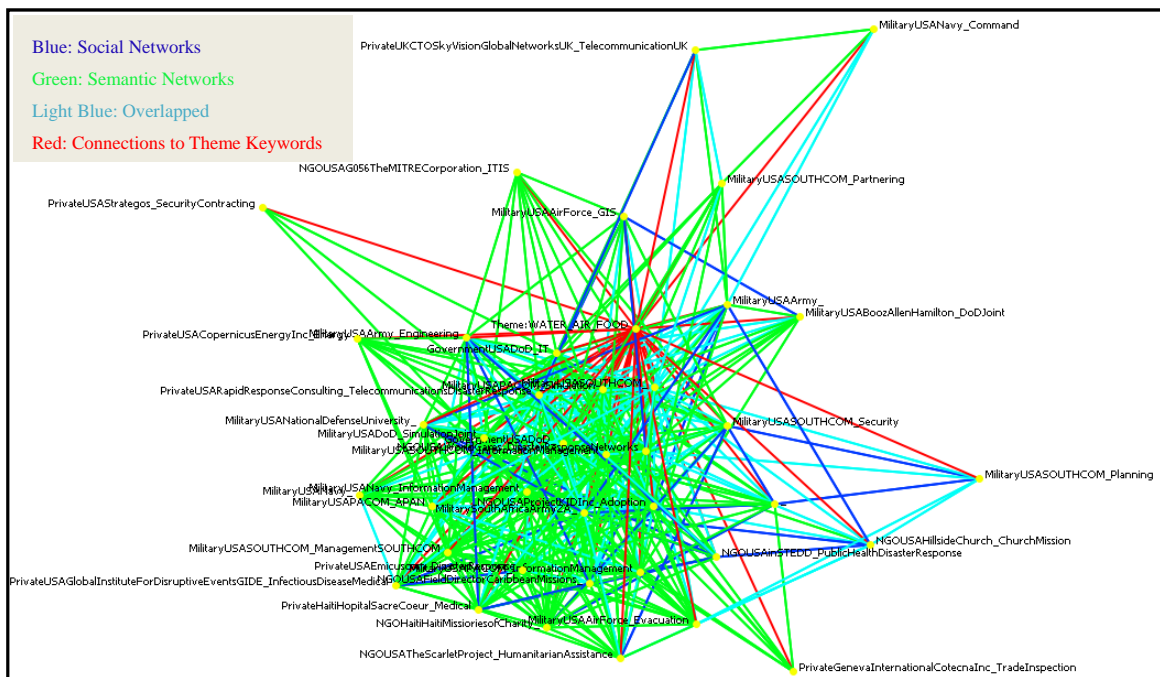


Figure 16: A holistic view of social network and semantic networks for the organizations associated with the theme “WATER, AIR, FOOD”.

	Social(Blue and Light Blue)	Overlapped (Light Blue)	Semantic (Green)	Involvement (Red)	Synergy	Top 10 for the First or Second Week
HIGH,RATES	16	16	20	9	100%	Yes
PEOPLE,CHILDREN,POST	140	124	308	26	89%	Yes
APAN,CONTACT,USER	186	160	272	38	86%	Yes
EARTHQUAKE,LEFT,MASSIVE,QUAKE	58	48	184	20	83%	Yes
AIRCRAFT,GROUND,ASSIST,ORDERED,USNS_COMFORT	116	96	102	21	83%	Yes
SUPPORT,PERSONNEL,CLIENT,UNIT	90	74	128	23	82%	Yes
MEDICAL,TIME,LIMITED	116	93	170	26	80%	No
WATER,AIR,FOOD	264	206	511	36	78%	Yes
RELIEF,FIELD,EFFORT	186	142	386	32	76%	No
INFORMATION,COMMUNITY,REQUESTS	234	176	296	35	75%	No
PATIENTS,LONG,CARE	104	74	262	25	71%	Yes
HAITI,OFFICER,PRESIDENT	130	90	298	33	69%	Yes
PAP,EMERGENCY,ASSESSMENT,ASSISTANCE	252	174	312	37	69%	NO

Figure 17: Synergy measured using the ratios of overlapping connections in the social networks and semantic networks.

5 Conclusion

In this paper, we explored a real-life data set, i.e. APAN forum data in the Haiti interagency operation to illustrate using the Lexical Link Analysis 1) to discover the themes and sort them according to the importance and time; 2) to extract social and semantic networks of participating organizations.

In summary, the LLA method can be used as a tool to examine historic data and improve future operation; identify partners for collaboration or areas of overlap, for example, as a monitoring tool to visualize organizational behavior and examining the aspects of the situational awareness of an interagency operation in real-time. Automation can be achieved via computer programs or software agents such as CLA to perform this task frequently – and in near real-time, thus placing “sifted” data into the hands of analysts allowing them the opportunity to understand revealed themes and evolving trends contained within the data that heretofore had been buried and likely left poorly understood.

6 References

- AutoMap. (2009). AutoMap: Extract, Analyze and Represent Relational Data from Texts. Retrieved from <http://www.casos.cs.cmu.edu/projects/automap/>
- Blei, D., Ng, A., & Jordan, M. (2003). Latent *Dirichlet* allocation. *Journal of Machine Learning Research*, 3, 993–1022. Retrieved from <http://jmlr.csail.mit.edu/papers/volume3/blei03a/blei03a.pdf>
- Dumais, S. T., Furnas, G. W., Landauer, T. K., Deerwester, S., & Harshman, R. (1988). Using latent semantic analysis to improve information retrieval. In *Proceedings of CHI'88: Conference on Human Factors in Computing* (pp. 281–285). New York, NY: Association for Computing Machinery.
- Girvan, M., & Newman, M. E. J. (2002, June). Community structure in social and biological networks. In the *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 99, No. 12,, p. 7821-7826.
- Gerber, C. (2005). Smart searching, new technology is helping defense intelligence analysts sort through huge volumes of data. *Military Information Technology*, 9(9). Retrived from <http://www.mkbergman.com/171/large-scale-intelligence-analysis/>
- Gorman, J. C., Foltz, P. W. Kiekel, P. A., Martin, M. A. & Cooke, N. J. (2003) Evaluation of Latent Semantic Analysis-based measures of communications content. In *Proceedings of the 47th Annual Human Factors and Ergonomic Society Meeting*.

Letsche, T.A. & Berry, M.W. (1997). Large-scale information retrieval with Latent Semantic Indexing. In *Proceedings of Inf. Sci.* 105-137.

USAID (2010), USAID Haiti Earthquake Fact Sheet. Retrived from
<http://www.usaid.gov/helphaiti/documents/01.19.10-USAID-DCHAHaitiEarthquakeFactSheet7.pdf>

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge, England: Cambridge University Press.

Zhao, Y., Gallup, S.P., & MacKinnon, D.J. (2011). Lexical Link Analysis for the Haiti Earthquake Relief Operation Using Open Data Source. In *Proceedings of the 16th ICCRTS, International Command and Control, Research and Technology Symposium*, Québec City, Canada June 21–23, 2011.



Semantic and Social Networks Comparison for the Haiti Earthquake Relief Operations from APAN Data Sources using Lexical Link Analysis (LLA)

Dr. Ying Zhao, Research Associate Professor, Naval Postgraduate School

Dr. Douglas J. MacKinnon, Research Associate Professor, Naval Postgraduate School

Dr. Shelley P. Gallup, Research Associate Professor, Naval Postgraduate School

17th ICCRTS, International Command and Control, Research and Technology
Symposium, Fairfax, Virginia, June 19–21, 2012



Research Questions



- How can we discover themes and topics in the discussion forum and sort the importance of the themes?
- How can we discover social and semantic networks of organizations that were involved, compare the two networks to obtain new insights?
 - What were the organizations involved in the *important* themes in the Haiti operation?
 - How do semantic networks offer more potential collaboration when compared to social networks?



HAITI Project

- Data Sources

- ~2600 open source web pages

- Used High Performance Computer (HPC) nodes.
 - News feed from 1/13 – 2/23/2010
 - Twitter as a starting point
 - SOUTHCOM, USAID and others used Twitter to handle the situations

- US SOUTHERN COMMAND also engaged the HAITI HA/DR Community of Interest (COI) on the All Partners Access Network (APAN) during the Haiti crisis. The APAN data was captured from a MSSQL database. The sources were:

- Official documents and briefings in PDF: ~167 PDF file attachments related to HAITI HA/DR from 1/13/2010 to 5/26/2010
 - SITREP: ~150 Situation Report documents
 - Forum: ~1173 posts from 1/13/2010 to 6/3/2010
 - Blogs: ~3900 blog messages

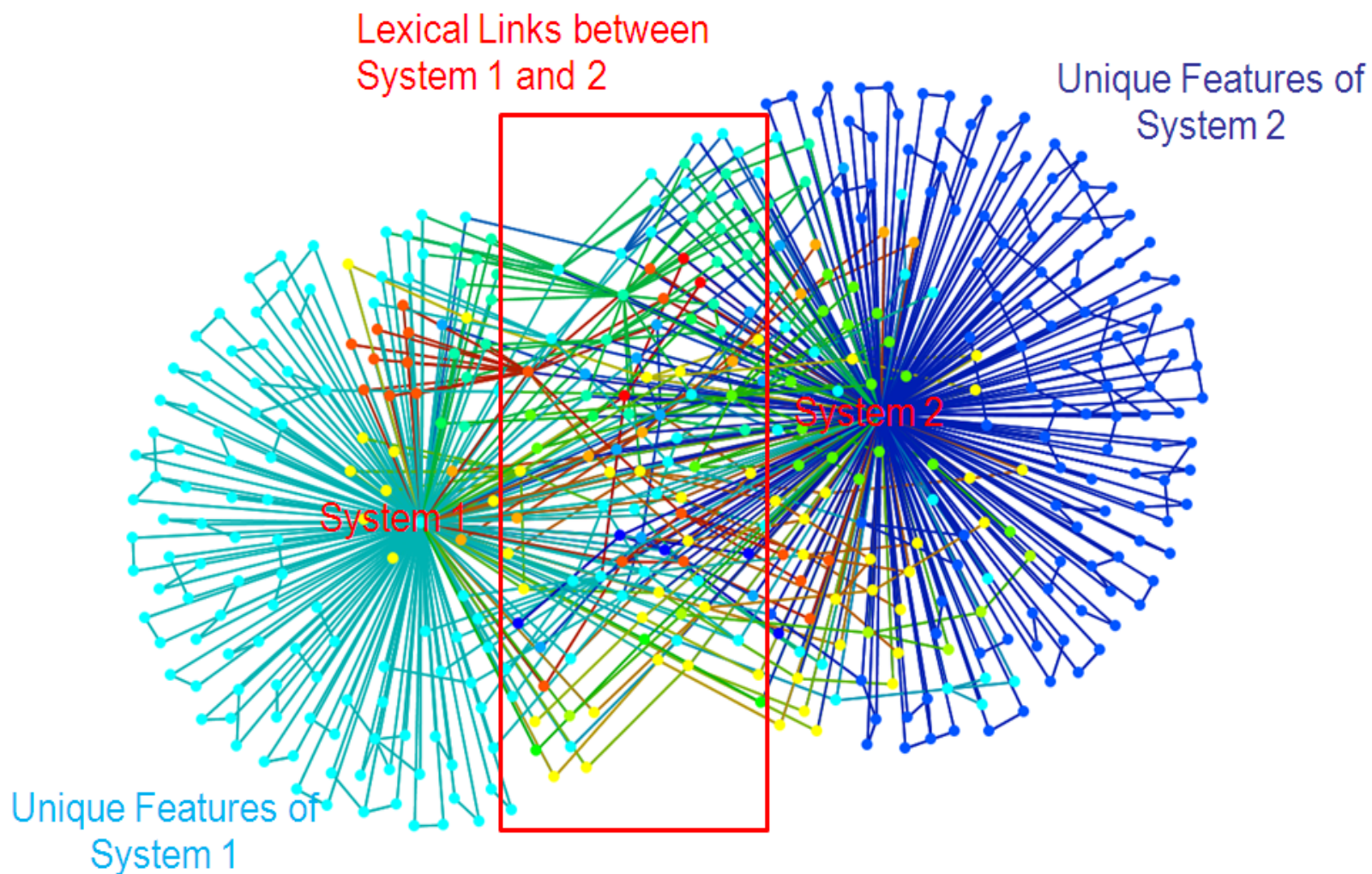


Compare Approaches

- Linguistics based
 - Part-of-speeches (e.g. nouns, verbs): SAP Business Objects/InXight NEE
 - Linguistics rules
- Statistical Co-occurrence
 - Tf*idf (http://en.wikipedia.org/wiki/Tf*idf)
 - Bi-gram, N-gram
- Representation
 - Bag-of-words (BOW), Latent Semantic Analysis (LSA), Latent Dirichlet Allocation (LDA), Support Vector Machines (SVM, accurate on text categorization, cloud/large-scale implementation example like yottamine)
 - Text-as-networks (TAN): Centering Resonance Analysis (CRA), Stanford Lexical Parser (syntactic dependence), AutoMap ,Google PageRank (DAN, Document-as-Networks),



LLA



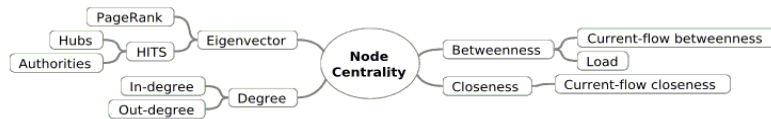


Why LLA?

Term Evaluation

Once the document is constructed as a network, the next step is to evaluate the nodes to determine the importance of each term. To do this, we employ the concept of *node centrality*. The centrality of a node is a measure of how important it is compared to the other nodes in the network.

There exists a myriad of different node centrality measures. These can roughly be divided into four main categories as illustrated by the figure below.



The simplest category is the degree-based centrality. These define node centrality in terms of the number and strengths of connections between a node and its neighbors.

The eigenvector centralities are also based on the node's neighbours, but is a bit more complex. These measures capture not only the number of neighbors a node has, but also take into account the importance of each neighbour. This group include among others, Google's famous PageRank algorithm.

The groups of closeness and betweenness centrality measures are a bit different, in that they focus more on the overall network structure. Closeness centralities are defined in terms of the lengths of the shortest paths from a node to the rest of the nodes in the network. Betweenness centrality describes whether, and how frequently, a node is part of the shortest paths between pairs of other nodes in the network.

In an attempt to find a node centrality measure well suited for capturing term importance, we experimented with all the measures shown in the figure above. The figure below shows classification accuracies when using representations based on different centrality measures.

- Critical needs
 - Links among documents/categories of information do not usually exist
 - Variety of network centrality measures can be used for term evaluations
- LLA automatically discovers relations among categories of text information to facilitate search, prioritization, prediction and expertise location
 - Generate hyperlinks for intranet documents
- LLA can also focus on **innovations and uniqueness** of the analyzed documents
 - Other ranking techniques which typically sort documents based on the *popularity* or *authority*, are not based on semantics
 - E.g. PageRank by Google



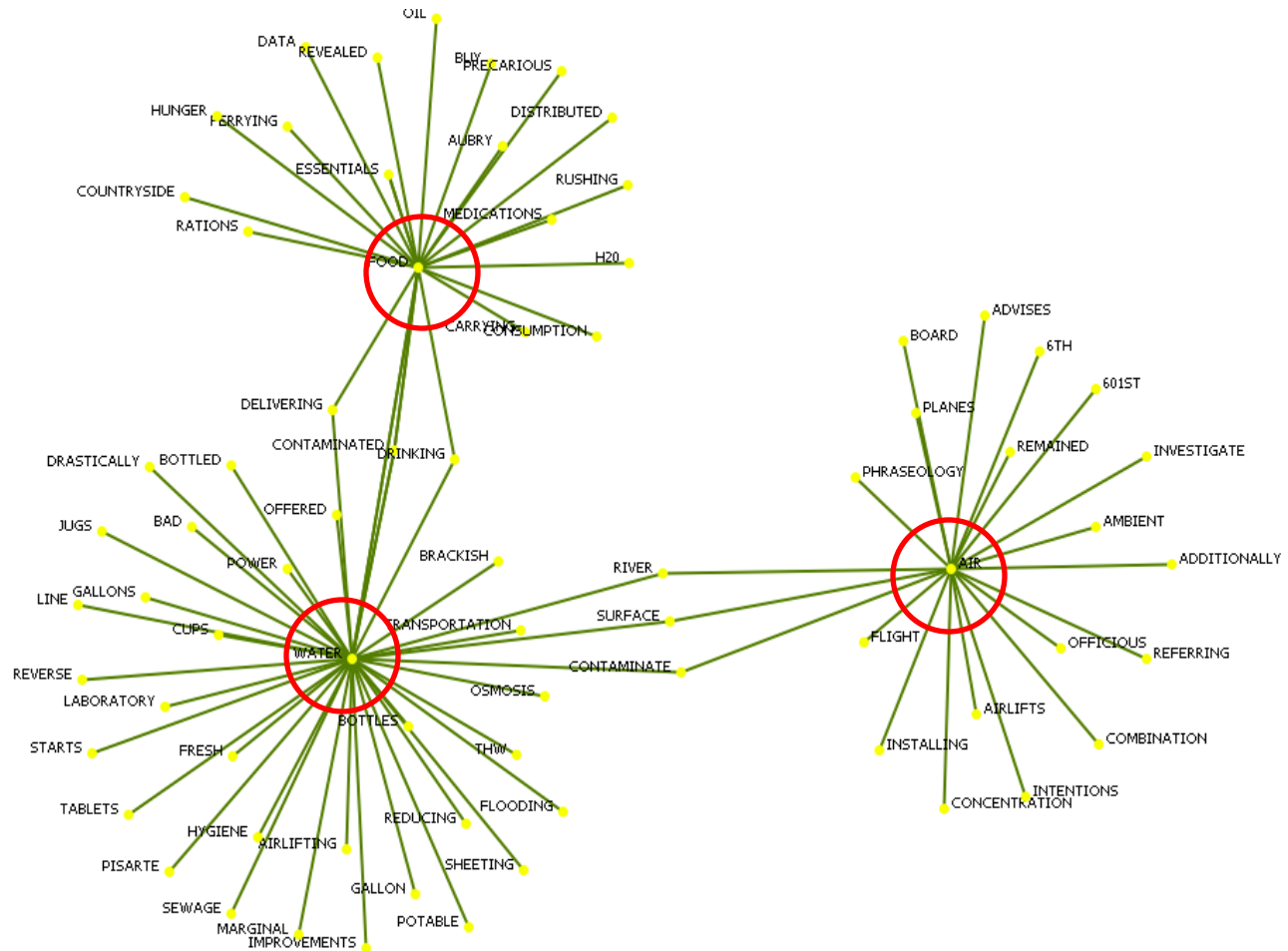
Discovered Themes



Event_Date_Sort	Theme Id	LLA Counts	Theme Keywords
ALL	95	83	WATER,AIR,FOOD
ALL	65	67	PEOPLE,CHILDREN,POST
ALL	6	65	APAN,CONTACT,USER
ALL	people	65	HAITI,OFFICER,PRESIDENT
ALL	64	60	PAP,EMERGENCY,ASSESSMENT,ASSISTANCE
ALL	41	58	INFORMATION,COMMUNITY,REQUESTS
ALL	54	56	MEDICAL,TIME,LIMITED
ALL	29	56	EARTHQUAKE,LEFT,MASSIVE,QUAKE
ALL	46	51	PATIENTS,LONG,CARE
ALL	70	50	RELIEF,FIELD,EFFORT
ALL	73	50	GROUP,SEE,FRENCH
ALL	96	50	DE,POSTED,INITIAL,WORKING
ALL	22	50	COUNTRY,HEALTH,AREAS,CHILE
ALL	place	49	STATE,PORT_AU_PRINCE,CITY
ALL	78	47	BUILDING,HQ,CIVIL,USERS
ALL	86	44	TEAM,COORDINATION,LEAD
ALL	39	42	HOSPITAL,BED,HOUSE
ALL	82	42	SUPPORT,PERSONNEL,CLIENT,UNIT



A “Theme” Defined



A collections of word pairs for the theme associated with keywords “WTER, AIR, FOOD”



Distributions of LLA Counts per Week



Dates	Theme ID	LLA Counts	Theme Keywords
ALL	95	83	WATER,AIR,FOOD
1/20/2010	95	39	WATER,AIR,FOOD
1/13/2010	95	37	WATER,AIR,FOOD
1/27/2010	95	20	WATER,AIR,FOOD
3/10/2010	95	7	WATER,AIR,FOOD
2/10/2010	95	3	WATER,AIR,FOOD
2/17/2010	95	3	WATER,AIR,FOOD
2/24/2010	95	3	WATER,AIR,FOOD
2/3/2010	95	2	WATER,AIR,FOOD
3/3/2010	95	2	WATER,AIR,FOOD
ALL	65	67	PEOPLE,CHILDREN,POST
1/20/2010	65	35	PEOPLE,CHILDREN,POST
1/13/2010	65	19	PEOPLE,CHILDREN,POST
1/27/2010	65	8	PEOPLE,CHILDREN,POST
2/10/2010	65	7	PEOPLE,CHILDREN,POST
3/3/2010	65	4	PEOPLE,CHILDREN,POST
2/24/2010	65	3	PEOPLE,CHILDREN,POST
2/3/2010	65	2	PEOPLE,CHILDREN,POST
2/17/2010	65	1	PEOPLE,CHILDREN,POST
3/10/2010	65	1	PEOPLE,CHILDREN,POST
3/21/2010	65	1	PEOPLE,CHILDREN,POST



Themes sorted for the first week (1/13/2010 to 1/19/201) and the second week (1/20/2010 to 1/26/2010)

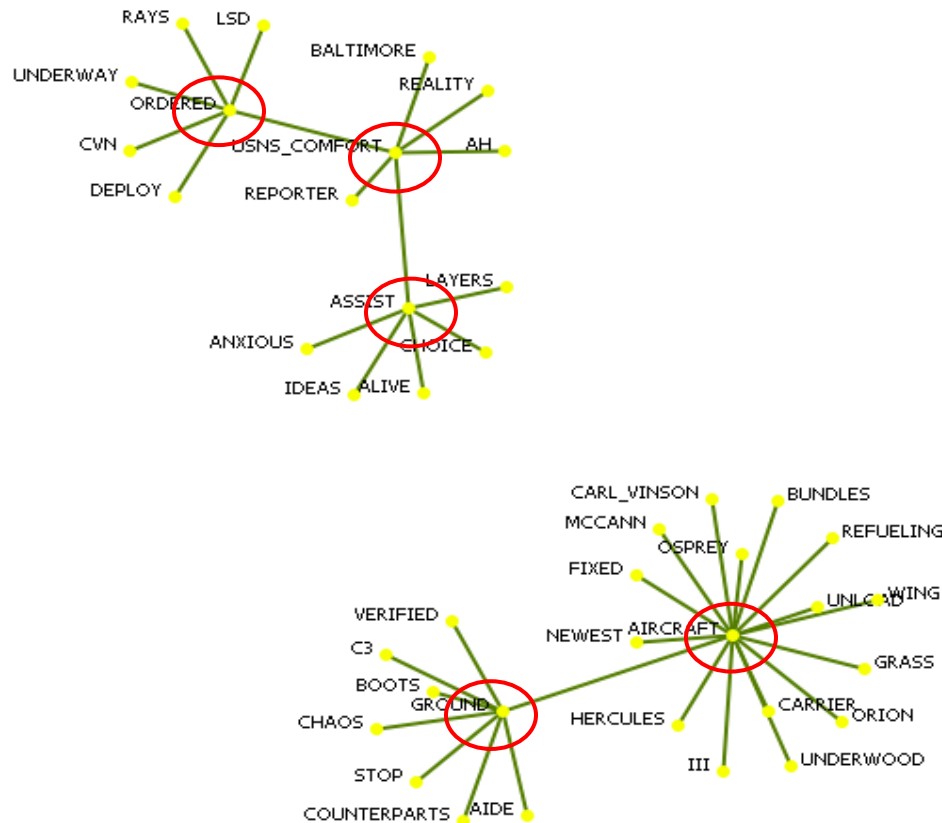
Event_Date_Sort	Theme Id	LLA Counts	Theme Keywords
1/13/2010	ALL	630	
1/13/2010	95	37	WATER,AIR,FOOD
1/13/2010	37	25	HIGH,RATES
1/13/2010	5	22	AIRCRAFT,GROUND,ASSIST,ORDERED,USNS_COMFORT
1/13/2010	6	21	APAN,CONTACT,USER
1/13/2010	82	21	SUPPORT,PERSONNEL,CLIENT,UNIT
1/13/2010	65	19	PEOPLE,CHILDREN,POST
1/13/2010	people	18	HAITI,OFFICER,PRESIDENT
1/13/2010	64	18	PAP,EMERGENCY,ASSESSMENT,ASSISTANCE
1/13/2010	79	18	STAFF,VERSION,BATTLE,LOGS,RECORDS
1/13/2010	69	18	PROVIDE,DEAD,NEEDED
1/13/2010	39	17	HOSPITAL,BED,HOUSE

Event_Date_Sort	Theme Id	LLA Counts	Theme Keywords
1/20/2010	95	39	WATER,AIR,FOOD
1/20/2010	65	35	PEOPLE,CHILDREN,POST
1/20/2010	people	33	HAITI,OFFICER,PRESIDENT
1/20/2010	46	32	PATIENTS,LONG,CARE
1/20/2010	29	26	EARTHQUAKE,LEFT,MASSIVE,QUAKE
1/20/2010	54	25	MEDICAL,TIME,LIMITED
1/20/2010	41	24	INFORMATION,COMMUNITY,REQUESTS
1/20/2010	6	23	APAN,CONTACT,USER
1/20/2010	70	23	RELIEF,FIELD,EFFORT
1/20/2010	72	23	RESPONSE,HELICOPTER,PART

- Important issues were heavily discussed and collaborated in the time of frame of two weeks.
 - The theme “HIGH, RATES” was related to the fact that the US government provided the analysis and information about high fatalities, high risks and high disease rates that might be expected after the event.



Lexical links for the theme associated with keywords “AIRCRAFT, GROUND, ASSIST, ORDERED, USNS_COMFORT”



- Measuring *like-mindedness*. Like-minded organizations do use *many* of the same terminologies in their communications. *like-mindedness* is closer in meaning to *like-wordedness*
- LLA often places the different terminologies such as “CVN” and “Carl_Vinson” in different components of word hubs but grouped together



The overlapping of top ten themes sorted by the LLA counts overall (ALL) and the first and second weeks

	Theme Keywords		Theme Keywords
All	WATER,AIR,FOOD	Week 1	WATER,AIR,FOOD
	PEOPLE,CHILDREN,POST		HIGH,RATES
	APAN,CONTACT,USER		AIRCRAFT,GROUND,ASSIST,ORDERED,USNS_COMFORT
	HAITI,OFFICER,PRESIDENT		APAN,CONTACT,USER
	PAP,EMERGENCY,ASSESSMENT,ASSISTANCE		SUPPORT,PERSONNEL,CLIENT,UNIT
	INFORMATION,COMMUNITY,REQUESTS	Week 2	WATER,AIR,FOOD
	MEDICAL,TIME,LIMITED		PEOPLE,CHILDREN,POST
	EARTHQUAKE,LEFT,MASSIVE,QUAKE		HAITI,OFFICER,PRESIDENT
	PATIENTS,LONG,CARE		PATIENTS,LONG,CARE
	RELIEF,FIELD,EFFORT		EARTHQUAKE,LEFT,MASSIVE,QUAKE

The LLA counts are measures to sort out the main and important topics and themes

- 6 out of 10 top themes (60%) in the top themes overall are also in the top themes of the first two weeks.
- 7 of 10 top themes in the first two weeks are also in the top overall themes



Extracting Social Networks

1	Thread	Level	Time	User	Organization	Content
43	5722	2	1/14/2010	2205	DoD U S Southern	re: attention!!! notice to airmen notam " That is correct. In order to gain aerial access
44	5722	2	1/15/2010	3342	Department of	re: attention!!! notice to airmen notam " I suggest that MTPP be used as a tag for all
45	5729	1	1/14/2010	3645	DoD Air Force	pap ground support I noticed this morning that there was a 60K main deck loader at
46	5736	1	1/14/2010	2205	DoD U S Southern	port-au-prince harbor port damage photos " [] [] [] [] [] "
47	5737	1	1/14/2010	2205	DoD U S Southern	rfi on behalf of mlgp haiti " Is there any info on a USCG helo flight inbound from Santo
48	5737	2	1/14/2010	2112	DoD US Southern	rfi on behalf of mlgp haiti " Heather, dont know, but we should send invite using the
49	5737	3	1/14/2010	2205	DoD U S Southern	rfi on behalf of mlgp haiti Yes! Lets do it. can you get them the directions or do you
50	5737	4	1/15/2010	2112	DoD US Southern	rfi on behalf of mlgp haiti " Heather, I dont want to step out of my lane. I think it
51	5741	1	1/14/2010	3645	DoD Air Force	mtpp airfield sitrep Can anyone provide any information regarding the Airfield
52	5741	2	1/15/2010	3342	Department of	mtpp airfield sitrep " I suggest that MTPP be used as a tag for all NOTAMs, blog posts,
53	5741	2	1/14/2010	3545	SSSG	mtpp airfield sitrep " Southern Command Press Release, QUOTE: "airfield has power
54	5741	3	1/15/2010	4038	Telecommunicatio	mtpp airfield sitrep " Both should be used, since Civilian Operators use NOTAMs as
55	5741	4	1/15/2010	3342	Department of	mtpp airfield sitrep Good point.
56	5743	1	1/14/2010	2406	Integrated ICT	usar activity as of
57	5744	1	1/14/2010	2406	Integrated ICT	logistical activity
58	5745	1	1/14/2010	2406	Integrated ICT	coordination act
59	5746	1	1/14/2010	2406	Integrated ICT	situation activity

List of threads of discussions

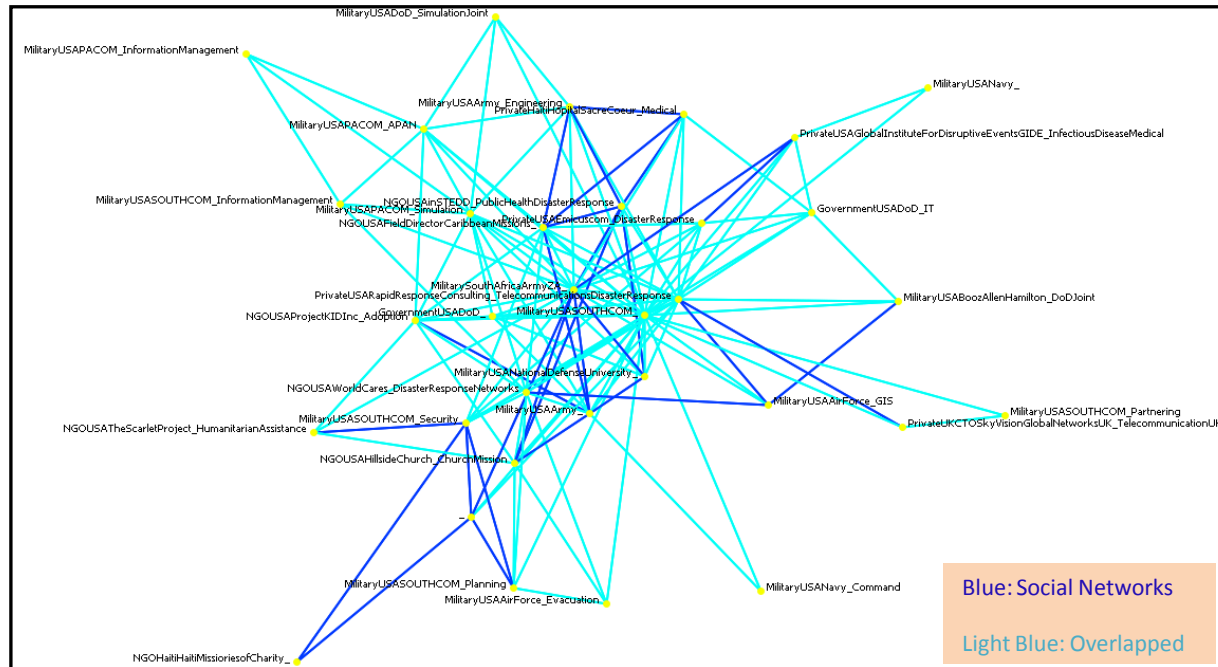
1	10022	10023	10033	10100	10113	10393	10406	10432	10473	10484	10491
78 MilitaryUSANORTHCOM_	0	0	0	0	0	0	0	0	0	0	0
79 MilitaryUSANORTHCOM_Joint	0	0	0	0	0	0	0	0	0	0	0
80 MilitaryUSANationalDefenseUniversity_	0	0	0	0	0	0	0	0	0	0	0
81 MilitaryUSANavy_	0	0	0	0	0	0	0	0	0	0	0
82 MilitaryUSANavy_Command	0	0	0	0	0	0	0	0	0	0	0
83 MilitaryUSANavy_InformationManagement	0	0	0	0	0	0	0	0	0	0	0
84 MilitaryUSANavy_Infrastructure	0	0	0	0	0	0	0	0	0	0	0
85 MilitaryUSANavy_Intelligence	0	0	0	0	0	0	0	0	0	0	0
86 MilitaryUSANavy_Joint	0	0	0	0	0	0	0	0	0	0	0
87 MilitaryUSANavy_JointIT	0	0	0	0	0	0	0	0	0	0	0
88 MilitaryUSANavy_LanguageInstitute	0	0	0	0	0	0	0	0	0	0	0
89 MilitaryUSANavy_Planning	0	0	0	0	0	0	0	0	0	0	0
90 MilitaryUSAPACOM_	0	0	0	0	0	0	0	0	0	0	0
91 MilitaryUSAPACOM_APAN	0	0	0	1	0	0	0	0	0	0	0
92 MilitaryUSAPACOM_InformationManagement	1	0	0	0	0	0	0	0	0	0	0
93 MilitaryUSAPACOM_Joint	0	0	0	0	0	0	0	0	0	0	0
94 MilitaryUSAPACOM_Planning	0	0	0	0	0	0	0	0	0	0	0
95 MilitaryUSAPACOM_Simulation	0	0	0	2	0	0	0	0	0	0	2
96 MilitaryUSASOUTHCOM_	0	0	0	0	0	0	1	1	0	0	0
97 MilitaryUSASOUTHCOM_Command	0	0	0	0	0	0	0	0	0	0	0
98 MilitaryUSASOUTHCOM_GIS	0	0	0	0	0	0	0	0	0	0	0
99 MilitaryUSASOUTHCOM_InformationManagement	0	0	0	0	0	0	0	0	0	0	0
100 MilitaryUSASOUTHCOM_Intelligence	0	0	0	0	0	0	0	0	0	0	0
101 MilitaryUSASOUTHCOM_Joint	0	0	0	0	0	0	0	0	0	0	0
102 MilitaryUSASOUTHCOM_ManagementSOUTHCOM	0	0	0	0	0	0	0	0	0	0	0
103 MilitaryUSASOUTHCOM_Partnering	0	0	0	0	0	0	0	0	0	0	0
104 MilitaryUSASOUTHCOM_Planning	0	0	0	0	0	0	0	0	0	0	0
105 MilitaryUSASOUTHCOM_Security	0	0	0	0	0	0	0	0	0	0	0
106 MilitaryUSASOUTHCOM_SpecialOperations	0	0	0	0	0	0	0	0	0	0	0

Organizations

Threads



Social Networks Associated with the Theme “WATER, AIR, DATA”



- Light blue links are social network links that also overlap in the semantic networks - indications of synergy



Extract Semantic Networks

```
PrivateUSACopernicusEnergyInc_Energy MilitaryUSAArmy_Engineering 5 ;INFORMATION--(189.00)
PrivateUSACopernicusEnergyInc_Energy MilitarySouthAfricaArmyZA_ 5 ;INFORMATION--(189.00)
PrivateUSACopernicusEnergyInc_Energy MilitaryUSASOUTHCOM_ 5 ;INFORMATION--(189.00)
PrivateUSACopernicusEnergyInc_Energy MilitaryCanadaNationalDefenseCA_ 5 ;INFORMATION--(189.00)
PrivateUSACopernicusEnergyInc_Energy MilitaryUSAPACOM_Simulation 5 ;INFORMATION--COMMUNITY(189.00)
PrivateUSACopernicusEnergyInc_Energy GovernmentUSADoD_ 5 ;SMALL--(208.00)
PrivateUSACopernicusEnergyInc_Energy MilitaryUSASOUTHCOM_ManagementSOUTHCOM 5 ;INFORMATION--(189.00)
PrivateUSACopernicusEnergyInc_Energy PrivateUSARapidResponseConsulting_TelecommunicationsDisasterResponse 5 ;INFORMATION--(189.00)
PrivateUSACopernicusEnergyInc_Energy MilitaryUSANavy_InformationManagement 5 ;POWER--(302.00)
PrivateUSACopernicusEnergyInc_Energy NGOUSAWorldCares_DisasterResponseNetworks 5 ;INFORMATION--(189.00)
PrivateUSACopernicusEnergyInc_Energy MilitaryUSADoD_SimulationJoint 5 ;SMALL--(208.00)
PrivateUSACopernicusEnergyInc_Energy MilitaryUSAPACOM_APAN 5 ;INFORMATION--REQUESTS(189.00)
PrivateUSACopernicusEnergyInc_Energy MilitaryUSADoD_IntelligenceJoint 5 ;INFORMATION--(189.00)
PrivateUSACopernicusEnergyInc_Energy PrivateUSAPaxMondial_Reconstruction 5 ;INFORMATION--CONFIDENTIAL(189.00)
MilitaryUSADoD_ITJoint MilitaryUSADoD_Informationsharing 5 ;SUSAN--(68.00)
MilitaryUSADoD_ITJoint MilitaryUSAEuropeanCommand_ 5 ;SUSAN--(68.00)
PrivateUSAEmicuscom_DisasterResponse MilitaryUSASOUTHCOM_Security 5 ;BOBBI--(1.00);JACMEL--(111.00)
PrivateUSAEmicuscom_DisasterResponse MilitaryUSAArmy_Engineering 5 ;BACKGROUND--(189.00)
PrivateUSAEmicuscom_DisasterResponse NGOUSAProjectKIDInc_Adoption 5 ;FOOD--(302.00);FOOD--WATER(302.00);JACMEL--(111.00)
PrivateUSAEmicuscom_DisasterResponse MilitarySouthAfricaArmyZA_ 5 ;FOOD--(302.00);REQUEST--(102.00)
PrivateUSAEmicuscom_DisasterResponse MilitaryUSASOUTHCOM_ 5
;GEAR--(97.00);FOOD--(302.00);GOOGLE--(225.00);REQUEST--(102.00);REQUEST--EXPLANATION(102.00);TRANSPORTING--(111.00);BACKGROUND--(189.00)
PrivateUSAEmicuscom_DisasterResponse MilitaryUSAPACOM_Simulation 5 ;FOOD--(302.00)
PrivateUSAEmicuscom_DisasterResponse GovernmentUSADoD_ 5
;FOOD--DRINKING,CONTAMINATED(302.00);FOOD--WATER(302.00);FOOD--WATER,DRINKING,CONTAMINATED(302.00);FOOD--(302.00)
PrivateUSAEmicuscom_DisasterResponse MilitaryUSASOUTHCOM_ManagementSOUTHCOM 5 ;FOOD--(302.00)
```

- Use shared word hubs



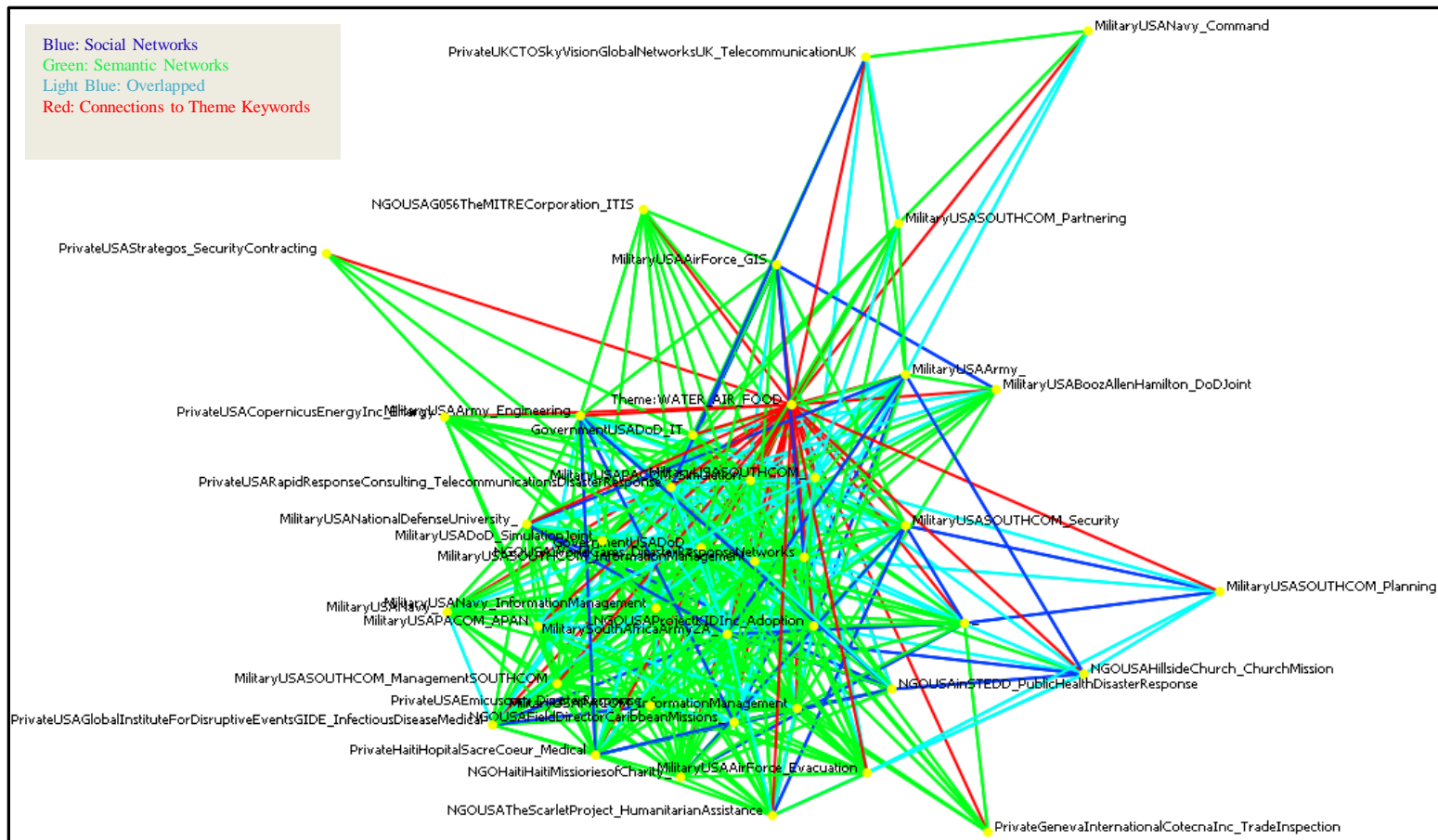
Semantic Networks (Green) Overlaid with the Social Networks for Theme “WATER, AIR, FOOD”



- Red circles show organizations who have semantic links but no current social connections => indicates potential collaborations



A holistic view of social network with semantic networks for the organizations associated with the theme “WATER, AIR, FOOD”





Measure Synergy Using Overlapping Connections in the Social networks and Semantic networks



Sorted according to the Synergy
Ratio=Overlapped (Light Blue)/Social (Blue and Light Blue)

	Social(Blue and Light Blue)	Overlapped (Light Blue)	Semantic (Green)	Involvement (Red)	Synergy	Top 10 for the First or Second Week
HIGH,RATES	16	16	20	9	100%	Yes
PEOPLE,CHILDREN,POST	140	124	308	26	89%	Yes
APAN,CONTACT,USER	186	160	272	38	86%	Yes
EARTHQUAKE,LEFT,MASSIVE,QUAKE	58	48	184	20	83%	Yes
AIRCRAFT,GROUND,ASSIST,ORDERED,USNS_COMFORT	116	96	102	21	83%	Yes
SUPPORT,PERSONNEL,CLIENT,UNIT	90	74	128	23	82%	Yes
MEDICAL,TIME,LIMITED	116	93	170	26	80%	No
WATER,AIR,FOOD	264	206	511	36	78%	Yes
RELIEF,FIELD,EFFORT	186	142	386	32	76%	No
INFORMATION,COMMUNITY,REQUESTS	234	176	296	35	75%	No
PATIENTS,LONG,CARE	104	74	262	25	71%	Yes
HAITI,OFFICER,PRESIDENT	130	90	298	33	69%	Yes
PAP,EMERGENCY,ASSESSMENT,ASSISTANCE	252	174	312	37	69%	NO

- There are fewer organizations involved in the two weeks than in the overall forum
- “Synergies” seem to go above the average (>80%) for some specific and time-critical themes, for example, “HIGH,RATES”, “PEOPLE,CHILDREN,POST”, “APAN,CONTACT,USER”, “AIRCRAFT,GROUND,ASSIST,ORDERED,USNS_COMFORT”, “SUPPORT,PERSONNEL,CLIENT,UNIT”, and “EARTHQUAKE,LEFT,MASSIVE,QUAKE”
- There are some themes with relatively low synergies overall and on the two weeks, for example, “WATER, AIR, FOOD”, “PATIENTS, LONG, CARE”, “HAITI, OFFICER, PRESIDENT”. This may indicate lack of knowledge or expertise in these areas in general.
- Comparing the two weeks and the average, the synergies among organizations seem correlated with the number of organizations involved in the collaborations: the themes with fewer organizations involved have higher synergies.



Conclusions

- Explored a real-life data set, i.e. APAN forum data in the Haiti interagency operation via Lexical Link Analysis
 - Discovered the themes and sorted them according to the importance and time
 - Extracted social and semantic networks of participating organizations
- Showed the LLA method to examine historic data and improve future operations
 - Identify partners for collaboration or areas of overlap
 - Examined the aspects of the situational awareness of an interagency operation
 - Monitored and visualized organizational behavior
 - Automation allows performing this task easier and more frequent – and in near real-time



DISE



Backup Slides



Lexical Link Analysis (LLA)



Summary

- Lexical Analysis (LA wiki, 2009) is a form of text mining
 - Learns
 - Dynamically updates word and context associations with added data
- Link analysis
 - Like network analysis, explores and illustrates associations between objects
- Lexical Link Analysis (LLA)
 - Combines data mining with network analysis
 - Can dynamically identify, assess, and predict trends, patterns and features
- Data mining tools
 - Analyzes structured and unstructured data
 - Confirms previously known patterns, or discovers new patterns
 - Implements innovative visualization and navigation techniques
- Facilitates concept discovery, automated classification, and categorization of unstructured documents

Agent Learning

